

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日 2 0 0 2 年 1 1 月 2 8 日
Date of Application:

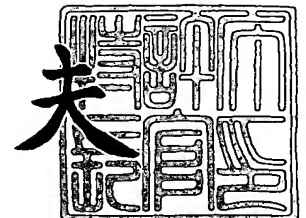
出 願 番 号 特 願 2 0 0 2 - 3 4 5 0 6 0
Application Number:
[ST. 10/C]: [J P 2 0 0 2 - 3 4 5 0 6 0]

出 願 人 沖 電 気 工 業 株 式 会 社
Applicant(s):

2 0 0 3 年 1 1 月 2 8 日

特許庁長官
Commissioner,
Japan Patent Office

今 井 康 夫



【書類名】 特許願

【整理番号】 MA001412

【提出日】 平成14年11月28日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 15/00

【発明者】

 【住所又は居所】 東京都港区虎ノ門1丁目7番12号 沖電気工業株式会社
社内

 【氏名】 佐々木 美樹

【発明者】

 【住所又は居所】 東京都港区虎ノ門1丁目7番12号 沖電気工業株式会社
社内

 【氏名】 池野 篤司

【特許出願人】

 【識別番号】 000000295

 【氏名又は名称】 沖電気工業株式会社

 【代表者】 篠塚 勝正

【代理人】

 【識別番号】 100083840

 【弁理士】

 【氏名又は名称】 前田 実

【選任した代理人】

 【識別番号】 100116964

 【弁理士】

 【氏名又は名称】 山形 洋一

【手数料の表示】

 【予納台帳番号】 007205

 【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9003703

【包括委任状番号】 0101807

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 数量表現検索装置

【特許請求の範囲】

【請求項 1】 自然言語で数量表現を検索する数量表現検索装置において、
検索対象の文書や検索したい数量表現を入力する入力手段と、
上記入力された文書や数量表現の文の構造を解析する文構造解析手段と、
属性を表す属性名と属性の意味を表す属性内容と補完するための基本単位とからなる属性情報、および省略されたと判断するための接頭語と接頭語の意味を表す倍数とからなる単位系情報が格納された属性辞書と、
属性を表す属性名と属性名を判断するための共起語とからなる情報が格納された共起語辞書と、
上記入力された文書や数量表現に対して、上記解析された文構造と上記属性辞書を参照してまたはさらに上記共起語辞書も参照して、上記文書や数量表現の接頭語に基本単位を補完することによって、省略された数量表現を補完する省略補完手段と
を備えた
ことを特徴とする数量表現検索装置。

【請求項 2】 請求項 1 記載の数量表現検索装置において、
上記補完後の文書から接頭語に基本単位を補完した語を検索キーワードとして抽出する抽出手段と、
文書データが格納される文書データベースと、
上記補完後の文書と上記入力された元の文書と上記抽出された検索キーワードとを上記文書データベースに格納する文書格納検索手段と
をさらに備え、

上記省略補完手段は、上記解析された文構造と上記共起語辞書を参照して上記入力された文書から接頭語のみに省略された数量表現を探し、その省略された数量表現に対して、上記解析された文構造をもとに上記接頭語の共起語を決定し、この決定した共起語をもとに上記共起語辞書を参照して上記接頭語の属性名を決定し、この決定した属性名をもとに上記属性辞書を参照して上記接頭語に基本単

位を補完する

ことを特徴とする数量表現検索装置。

【請求項 3】 請求項 2 記載の数量表現検索装置において、
出力手段をさらに備え、

上記省略補完手段は、上記解析された文構造と上記共起語辞書を参照して上記入力された数量表現が接頭語のみに省略された数量表現であるか否かを判別し、接頭語のみに省略された数量表現であれば、その旨を上記出力手段によって通知し、再入力を促す

ことを特徴とする数量表現検索装置。

【請求項 4】 請求項 2 記載の数量表現検索装置において、
出力手段をさらに備え、

上記省略補完手段は、上記解析された文構造と上記共起語辞書を参照して上記入力された数量表現が接頭語のみに省略された数量表現であるか否かを判別し、接頭語のみに省略された数量表現であれば、基本単位と属性情報を上記出力手段によって提示して、基本単位の内の 1 つを選択するように促し、選択された基本単位を用いて補完する

ことを特徴とする数量表現検索装置。

【請求項 5】 請求項 4 記載の数量表現検索装置において、
上記省略補完手段は、省略された数量表現を補完するための基本単位が選択されなかったときに、補完可能な全ての基本単位を用いて補完をする

ことを特徴とする数量表現検索装置。

【請求項 6】 請求項 3 から 5 までのいずれかに記載の数量表現検索装置において、

上記文書格納検索手段は、上記入力された数量表現、上記選択された基本単位で補完された数量表現、または上記補完可能な全ての基本単位で補完された数量表現と検索キーワードが一致する文書を上記文書データベースから検索し、上記検索結果として上記出力手段によって出力する

ことを特徴とする数量表現検索装置。

【発明の詳細な説明】

【0001】**【発明の属する技術分野】**

本発明は、自然言語で数量表現を検索する数量表現検索装置に関する。

【0002】**【従来の技術】**

自然言語で様々な表記された同じ意味の数量表現を検索できるようにするには、数量表現を変換する必要がある。

【0003】

従来の数量表現検索装置では、文書内をサーチし、数量表現テンプレートとマッチングを行うことによって、文書内の数量表現を一括して適切な数量表現に変換することが可能となり、機械翻訳システムなどに利用することができる（例えば、特許文献1参照）。

【0004】**【特許文献1】**

特開平5-67137号公報

【0005】**【発明が解決しようとする課題】**

しかしながら、上記従来の数量表現検索装置では、単語の意味情報と変換関数を用いて変換するだけなので、省略されていて意味が複数考えられる表現には対応することができない。

【0006】

例えば、「尺」を日本語の長さの数量表現に登録し、「centimeter」を英語の長さの数量表現に登録しておくと、「尺」を「centimeter」に変換することができるとなっているが、文書内で「キロ」と省略されている場合は、「キロメートル」なのか、「キログラム」なのか、判断できないので、正しく変換することはできない。

【0007】

本発明は、このような従来の従来の課題に鑑みてなされたものであり、接頭語にみに省略されている場合を意識しなくても数量表現を検索することができる数

量表現検索装置を提供することを目的とするものである。

【0008】

【課題を解決するための手段】

本発明の数量表現検索装置は、
検索対象の文書や検索したい数量表現を入力する入力手段と、
上記入力された文書や数量表現の文の構造を解析する文構造解析手段と、
属性を表す属性名と属性の意味を表す属性内容と補完するための基本単位とからなる属性情報、および省略されたと判断するための接頭語と接頭語の意味を表す倍数とからなる単位系情報が格納された属性辞書と、
属性を表す属性名と属性名を判断するための共起語とからなる情報が格納された共起語辞書と、

上記入力された文書や数量表現に対して、上記解析された文構造と上記属性辞書を参照してまたはさらに上記共起語辞書も参照して、上記文書や数量表現の接頭語に基本単位を補完することによって、省略された数量表現を補完する省略補完手段と

を備えた

ことを特徴とするものである。

【0009】

【発明の実施の形態】

図1は本発明の実施の形態の数量表現検索装置のブロック構成図である。この実施の形態の数量表現検索装置は、入力手段1と、文構造解析手段2と、省略補完手段3と、属性辞書4と、共起語辞書5と、文書格納検索手段6と、文書データベース7と、抽出手段8と、出力手段9とを備える。

【0010】

入力手段1は、検索対象の文書や検索したい数量表現を入力する手段である。この入力手段1は、入力された文書や数量表現を文構造解析手段2に送る。

【0011】

文構造解析手段2は、入力された文の構造を解析する手段である。この文構造解析手段2は、入力手段1から送られた文書や数量表現の文構造を形態素解析と

構文解析によって解析し、解析後の文構造を入力された元の文書や数量表現とともに省略補完手段 3 に送る。

【0012】

省略補完手段 3 は、接頭語のみに省略された（省略されて接頭語のみである）数量表現に基本単位を補完する手段である。この省略補完手段 3 は、文構造解析手段 2 から送られた文構造をもとに、属性辞書 4 と共起語辞書 5 を参照して、文書や数量表現の接頭語に基本単位を補完し、補完後の文書や数量表現を入力された元の文書や数量表現とともに抽出手段 8 に送る。

【0013】

図 2 は数量表現を含む文の文構造の解析例を示す図である。数量表現の修飾先の語を、その共起語とする。図 2（1），（2），（3）の「5 M」の共起語は「メモリ」になる。また、図 2（4）の「5 M」の共起語は「増設する」になる。

【0014】

属性辞書 4 は、属性の情報と単位系の情報を格納する辞書である。この属性辞書 4 は、属性情報が属性名と属性内容と基本単位からなり、単位系情報が接頭語と倍数と基本単位からなる。

【0015】

共起語辞書 5 は、省略を補完する共起語の情報を格納する辞書である。この共起語辞書 5 は、属性名と共起語からなる。

【0016】

図 3 は属性辞書 4 の構成例を示す図であり、図 4 は共起語辞書 5 の構成例を示す図である。

【0017】

文書格納検索手段 6 は、文書を格納および検索する手段である。この文書格納検索手段 6 は、抽出手段 8 から入力された補完後の文書や元の文書や検索キーワードを文書データベース 7 に格納し、抽出手段 8 から入力された補完後の数量表現が検索キーワードと一致する文書を文書データベース 7 から検索して出力手段 9 に送る。

【0018】

文書データベース7は、検索対象の文書や補完後の文書を格納するデータベースである。

【0019】

抽出手段8は、検索キーワードを抽出する手段である。この抽出手段8は、省略補完手段3から入力された補完後の文書や数量表現と、補完後の語を抽出した検索キーワードを、文書格納検索手段6に送る。

【0020】

出力手段9は、結果を出力する手段である。この出力手段9は、文書格納検索手段6から送られた検索結果を出力する。

【0021】

なお、形態素解析を行う処理、構文解析を行う処理、文書をデータベース化する処理、文書を格納または検索する処理、および該当部分（検索キーワード）を抽出する処理は、一般的な部分に関しては、既知の自然言語処理技術を用いればよい。

【0022】

図5は本発明の実施の形態の数量表現検索装置の動作を説明するフローチャートである。この図5において、入力手段1で処理を選択し（ステップ501）、投入処理を実行するか（ステップ502）、検索処理を実行するか（ステップ503）、または終了する。

【0023】

図6は図5のステップ502の投入処理の動作を説明するフローチャートである。

【0024】

図6の投入処理では、まず最初に、検索対象の文書を入力手段1に投入する（ステップ601）。

例えば、以下の例文（a）または（b）を投入する。

「10キロの荷物を持って歩いた」…（a）

「10キロ荷物を持って歩いた」…（b）

入力手段 1 に投入された文書は、文書解析手段 2 に送られる。

【0025】

次に、文構造解析手段 2 において、投入された文書の文構造を解析する（ステップ 602）。

図 7 は例文（a）、（b）の文構造の解析例であり、（1）は例文（a）の解析例、（2）は例文（b）の解析例である。

文構造解析手段 2 での解析後の文構造は、入力手段 1 から送られた元の文書とともに、省略補完手段 3 に送られる。

【0026】

次に、省略補完手段 3 において、解析後の文構造および属性辞書 4（図 3 参照）の単位系情報を参照して、文書から接頭語を探す（ステップ 603）。

例文（a）、（b）では、接頭語として、ともに「キロ」が探される。

【0027】

なお、上記ステップ 603 から以下のステップ 607 までの処理は、省略補完手段 3 においての処理である。

【0028】

次に、文構造解析手段 2 で解析された文構造から共起語を決定する（ステップ 604）。

例文（a）の共起語は「荷物」と決定される。

例文（b）の共起語は「歩く」と決定される。

【0029】

次に、共起語辞書 5（図 4 参照）を参照して、属性（属性名）を決定する（ステップ 605）。

例文（a）の属性名は「WEIGHT」と決定される。

例文（b）の属性名は「LENGTH」と決定される。

【0030】

さらに、属性辞書 4（図 3）を参照して、基本単位を決定する（ステップ 606）。

例文（a）では、属性が「WEIGHT」なので、基本単位は「グラム」と決定され

る。

例文 (b) では、属性が「LENGTH」なので、基本単位は「メートル」と決定される。

【0031】

そして、接頭語に基本単位を補完する (ステップ 607)。

例文 (a) では、接頭語「キロ」に基本単位「グラム」が補完され、「10 キログラムの荷物を持って歩いた」となる。

例文 (b) では、接頭語「キロ」に基本単位「メートル」が補完され、「10 キロメートル荷物を持って歩いた」となる。

補完後の文書は、元の文書とともに、抽出手段 8 に送られる。

【0032】

次に、抽出手段 8 において、補完後の語を検索キーワードとして抽出する (ステップ 608)。

例文 (a) では、「10 キログラム」がキーワードとして抽出される。

例文 (b) では、「10 キロメートル」がキーワードとして抽出される。

抽出されたキーワードは、元の文書とともに、文書格納検索手段 6 に送られる。

。

【0033】

最後に、文書格納検索手段 6 によって、文書データベース 7 に、元の文書と検索キーワードを格納し (ステップ 609)、投入処理を終了する。

例文 (a) では、元の文書「10 キロの荷物を持って歩いていた」とキーワード「10 キログラム」が文書データベース 7 に格納される。

例文 (b) では、元の文書「10 キロ荷物を持って歩いていた」とキーワード「10 キロメートル」が文書データベース 7 に格納される。

【0034】

図 8 は図 5 のステップ 503 の検索処理の動作を説明するフローチャートである。

【0035】

図 8 の検索処理では、まず最初に、検索したい数量表現を、検索語として入力

手段1に入力する（ステップ801）。

例えば、以下の例文（c）または（d）を検索語として入力する。

「10キロメートル」…（c）

「10キロ」…（d）

入力手段1に入力された数量表現（検索語）は、文書解析手段2に送られる。

【0036】

次に、文構造解析手段2において、検索語の文構造を解析する（ステップ802）。文構造解析手段2での解析後の文構造は、入力手段1から送られた数量表現（検索語）とともに、省略補完手段3に送られる。

【0037】

次に、省略補完手段3において、解析後の文構造および属性辞書4（図3参照）の単位系情報を参照して、検索語が接頭語であるか否か（接頭語のみに省略された数量表現であるか否か）を判別する（ステップ803）。

例文（c）では、接頭語でないと判別される。

例文（d）では、「キロ」が接頭語であると判別される。

接頭語でないと判別された場合は、文書格納検索手段6に送られる。

【0038】

上記ステップ803で接頭語でないと判別された場合は、文書格納検索手段6において、検索語と検索キーワードが一致する文書を文書データベース7に格納されている文書から検索して獲得する（ステップ804）。

例文（c）では、検索キーワードが「10キロメートル」である上記例文（b）の「10キロ荷物を持って歩いた」が、文書データベース7から検索されて獲得される。

【0039】

そして、上記ステップ804で獲得した文書を検索結果として出力手段8から出力する（ステップ805）。

例文（c）では、上記例文（b）の「10キロ荷物を持って歩いた」が検索結果として出力される。

【0040】

また、上記ステップ803で接頭語であると判別された場合は、省略補完手段3において、属性辞書4（図3参照）の属性情報を参照して、基本単位と属性内容の一覧を出力手段9で表示させて、省略された数量表現であることを通知する（ステップ811）。

【0041】

なお、上記ステップ811から以下のステップ815までの処理は、省略補完手段3においての処理である。

【0042】

そして、再入力するか否かを出力手段9で表示させ、再入力するか否かを問い合わせる（ステップ812）。

【0043】

上記ステップ812で、再入力することが選択された場合には、上記ステップ801の検索語の入力に戻る。

【0044】

また、上記ステップ812で、再入力しないことが選択された場合には、基本単位から選択するか否かを出力手段9で表示させ、基本単位から選択するか否かを問い合わせる（ステップ813）。

【0045】

上記ステップ813で、基本単位からの選択がなされた場合には、その選択された基本単位で接頭語（検索語）を補完する（ステップ814）。

例文（d）では、例えば基本単位「グラム」が選択され、その基本単位「グラム」で検索語「10キロ」が補完され、「10キログラム」となる。

補完された検索語は、文書格納検索手段6に送られる。

【0046】

そして、上記ステップ814で検索語に補完がなされた場合には、文書格納検索手段6において、検索語と検索キーワードが一致する文書を文書データベース7に格納されている文書から検索して獲得し（ステップ804）、この獲得した文書を検索結果として出力手段8から出力する（ステップ805）。

例文（d）では、検索キーワードが「10キログラム」である上記例文（a）

の「10キロの荷物を持って歩いた」が、文書データベース7から検索されて獲得され、検索結果として出力手段8から出力される。

【0047】

また、上記ステップ813で、基本単位からの選択がなされなかった場合には、全ての基本単位で接頭語（検索語）を補完する（ステップ815）。

例文（d）では、例えば全ての基本単位「メートル」、「グラム」、「バイト」、…で、検索語「10キロ」が補完され、「10キロメートル」、「10キログラム」、「10キロバイト」、…となる。

全ての基本単位で補完されたこれらの検索語は、文書格納検索手段6に送られる。

【0048】

そして、上記ステップ815で検索語に補完がなされた場合には、文書格納検索手段6において、補完された全ての検索語について、変換後の検索キーワードが一致する文書を文書データベース7に格納されている文書からそれぞれ検索して獲得し（ステップ804）、これら獲得した文書を検索結果として出力手段8から出力する（ステップ805）。例文（d）では、検索キーワードが「10キログラム」である上記例文（a）の「10キロの荷物を持って歩いた」、検索キーワードが「10キロメートル」である上記例文（b）の「10キロ荷物を持って歩いた」などが、文書データベース7から検索されて獲得され、検索結果として出力手段8から出力される。

【0049】

共起語として修飾関係や格関係になる語を抽出する方法や、抽出した語の関係を表すシンソーラスを作成する手法や、抽出した語の関係から訳し分けを行う手法はすでに存在するが、それらの手法の対象は被修飾語や格になる名詞や動詞であり、それらの手法を用いても、修飾になる数量表現の属性を決定することができない。

【0050】

以上のように本発明の実施の形態によれば、文構造を解析して共起語を決定して省略された数量表現を補完して格納しておく、あるいは省略された数量表現を

適切に補完する語のみを検索時に提供して検索することによって、省略された数量表現が検索対象文書と検索語のいずれに存在していても、省略されていることを意識せずに補完して検索できる。また、自然言語の検索に用いると、数量表現の検索が容易になる。

【0051】

なお、上記実施の形態では、数値と単位による数量表現のみを検索対象または検索語とする数量表現検索装置を説明したが、本発明は、その他の数量表現や数量表現以外を検索対象または検索語とする方法や装置と組み合わせて利用することも可能である。

【0052】

【発明の効果】

以上説明したように本発明によれば、接頭語にのみ省略されていることを意識せずに数量表現を補完して検索できるという効果がある。

【図面の簡単な説明】

【図1】 本発明の実施の形態の数量表現検索装置のブロック構成図である。

【図2】 数量表現を含む文の文構造の解析例を示す図ある。

【図3】 図1の属性辞書の構成例を示す図である。

【図4】 図1の共起語辞書の構成例を示す図である。

【図5】 図1の数量表現検索装置の動作を説明するフローチャートである。

【図6】 図5のステップ502の投入処理の動作を説明するフローチャートである。

【図7】 図6のステップ602においての文構造の解析例を示す図である。

【図8】 図5のステップ503の検索処理の動作を説明するフローチャートである。

【符号の説明】

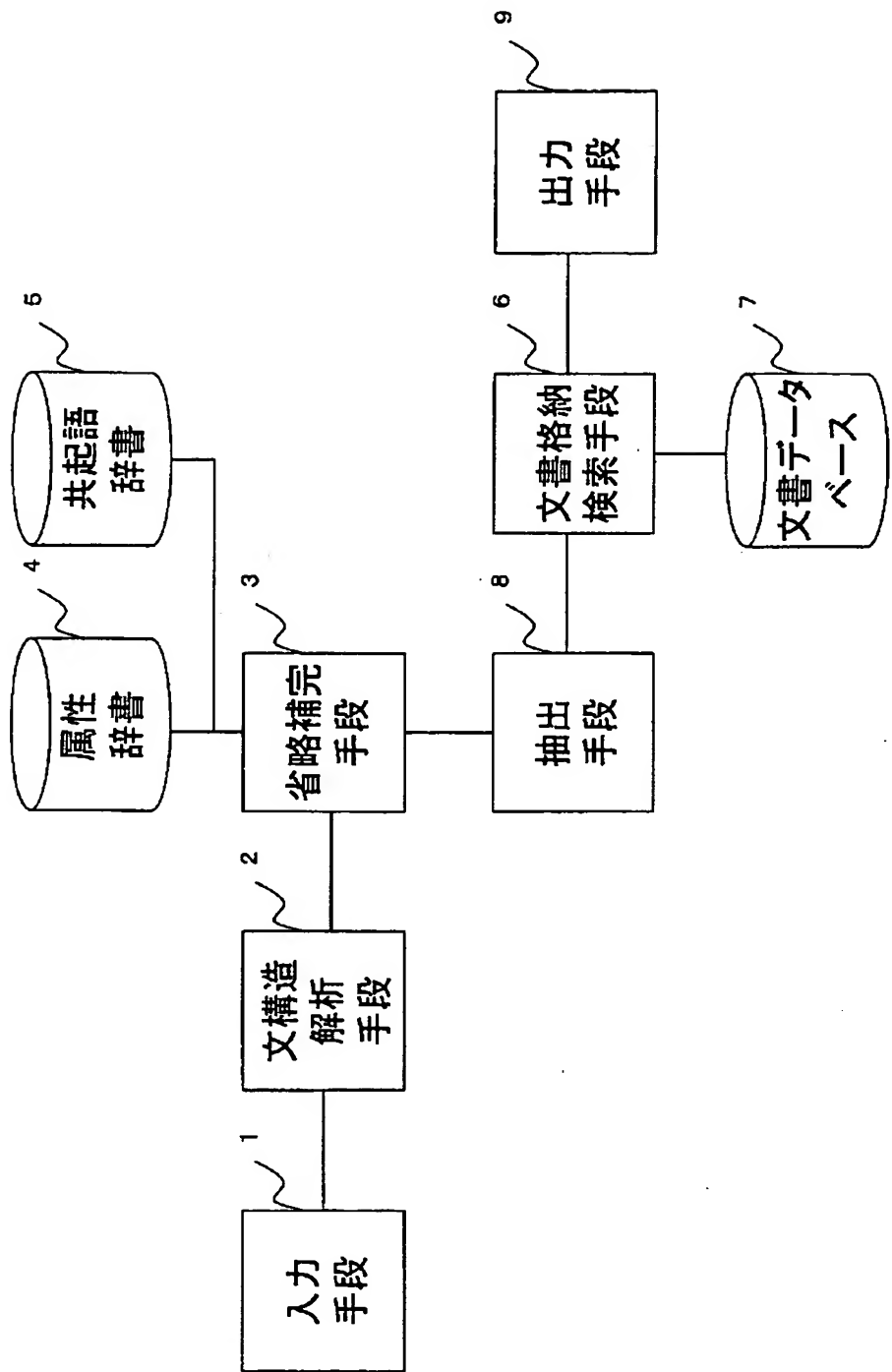
1 入力手段、 2 文構造解析手段、 3 省略補完手段、 4 属性辞書

、 5 共起語辞書、 6 文書格納検索手段、 7 文書データベース、 8
抽出手段、 9 出力手段。

【書類名】

図面

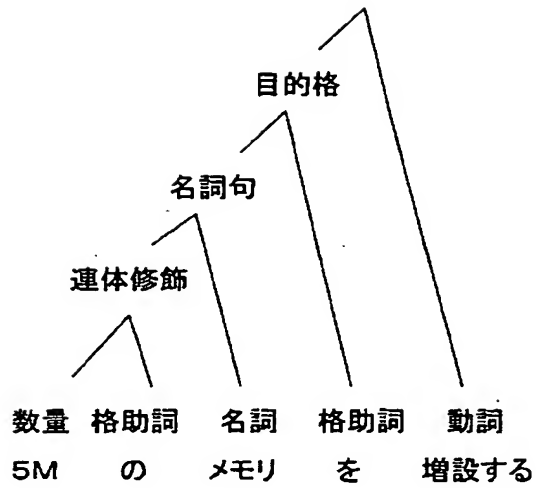
【図 1】



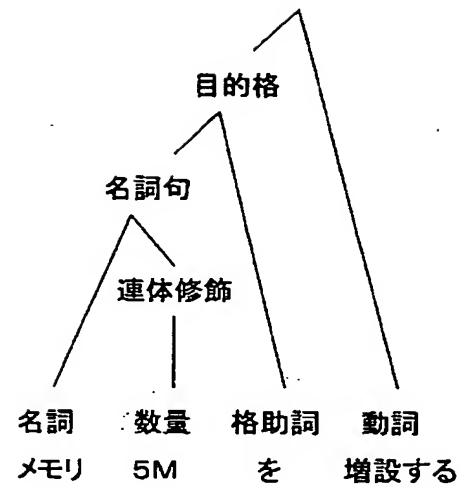
実施の形態の構成

【図 2】

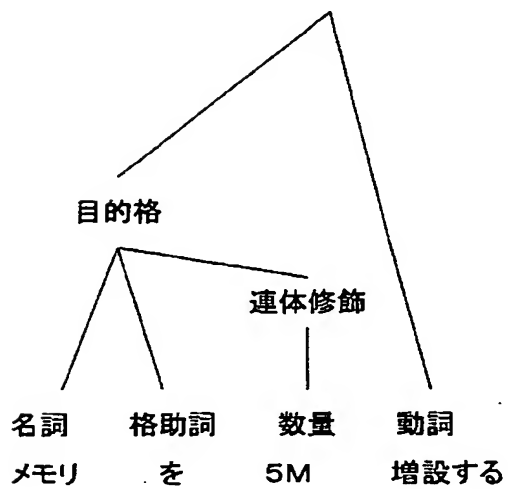
(1)



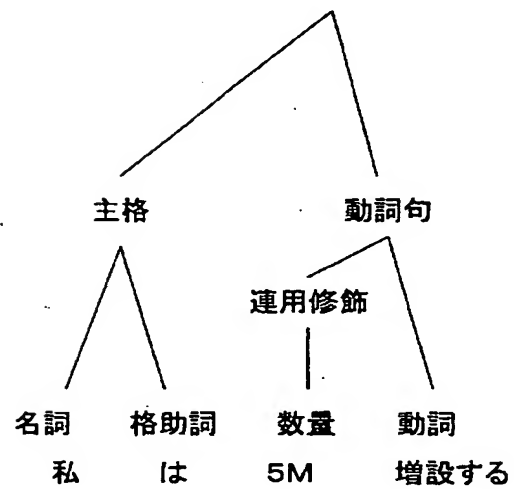
(2)



(3)



(4)



【図 3】

属性名	属性内容	基本単位	接頭語	倍数
LENGTH	長さ	メートル	ナノ、n	*10**-9
WEIGHT	重さ	グラム	キロ、k	*10**3
STORAGE	記憶容量	バイト	メガ、M	*10**6
...

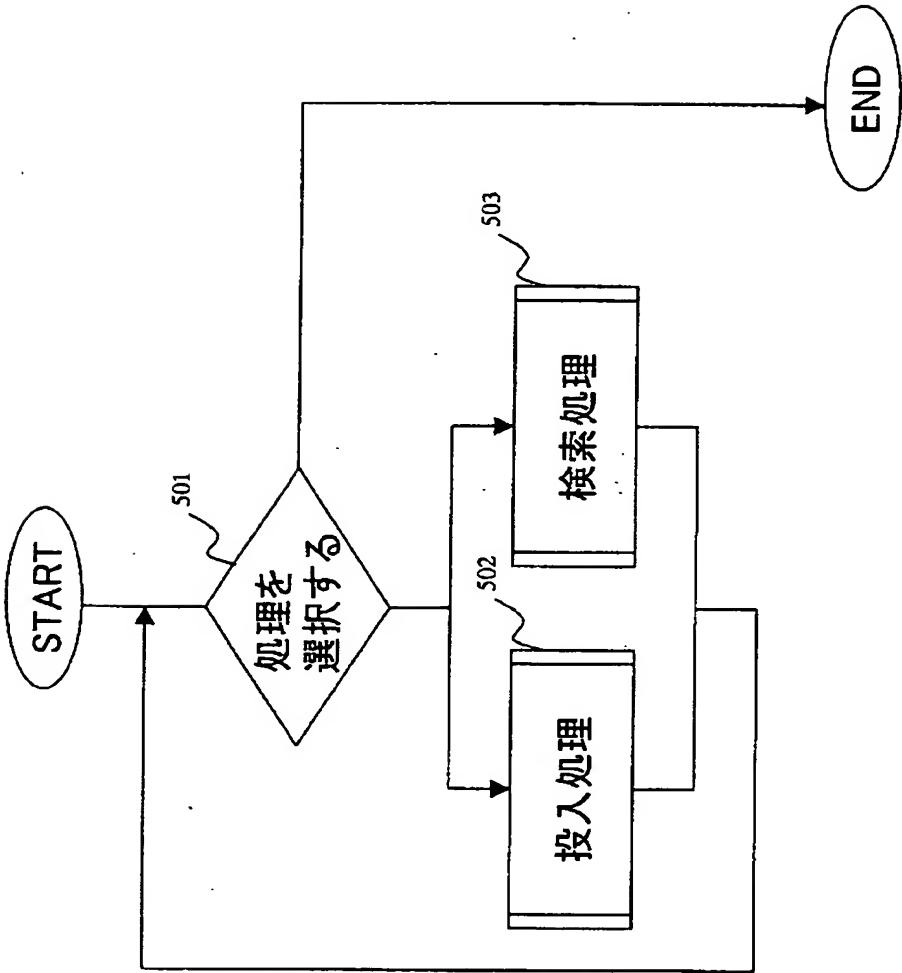
属性辞書の例

【図 4】

属性名	共起語
LENGTH	長さ、時速、距離、歩く、...
WEIGHT	重さ、荷物、持つ、...
STORAGE	記憶容量、メモリ、増設、...
...	...

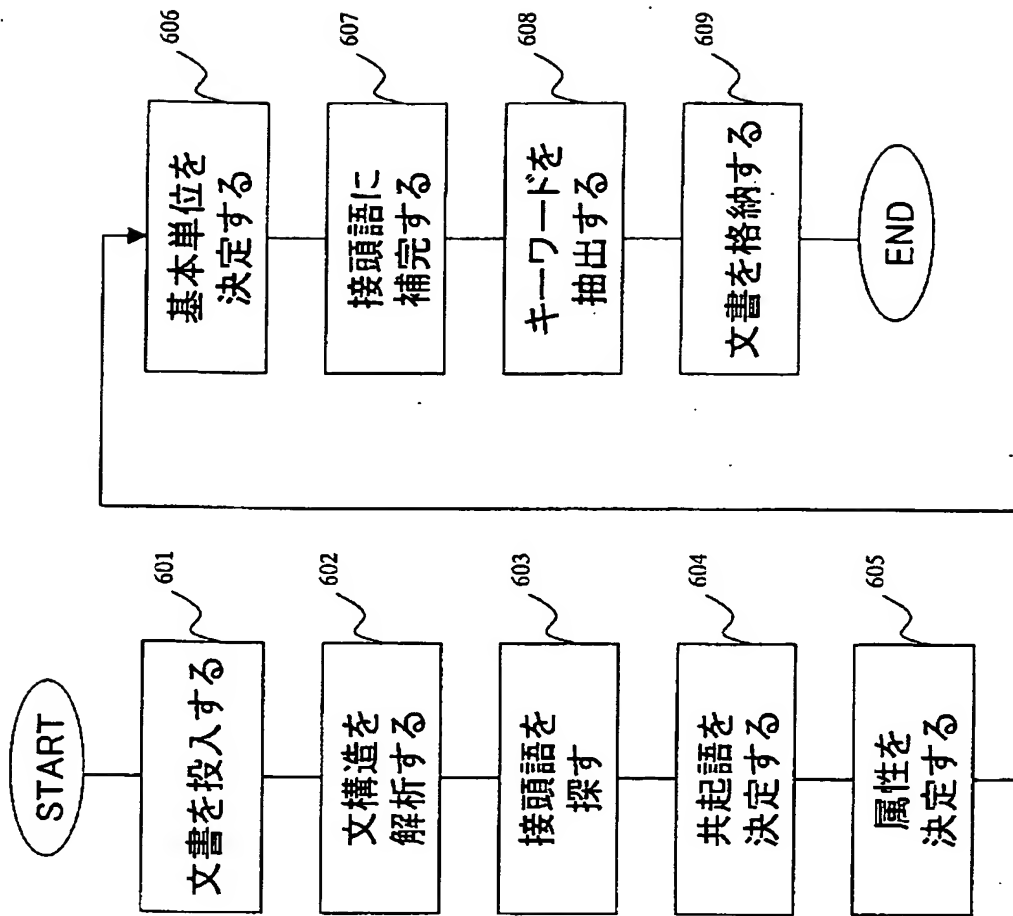
共起語辞書の例

【図 5】



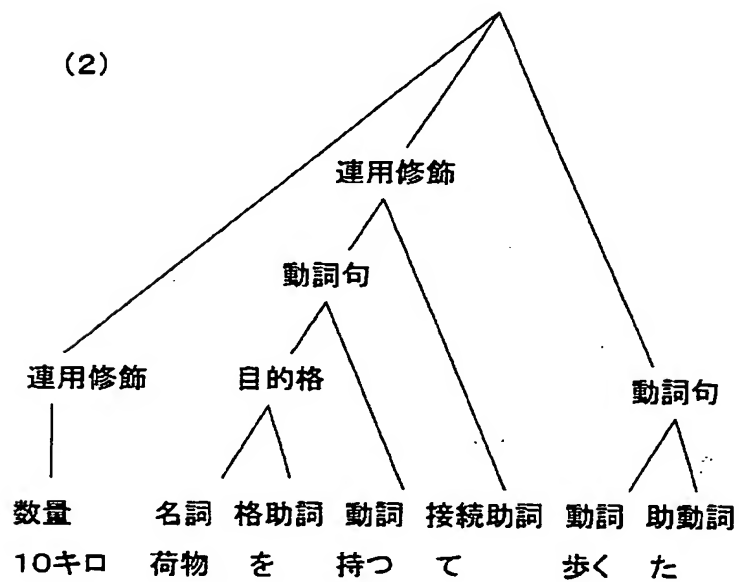
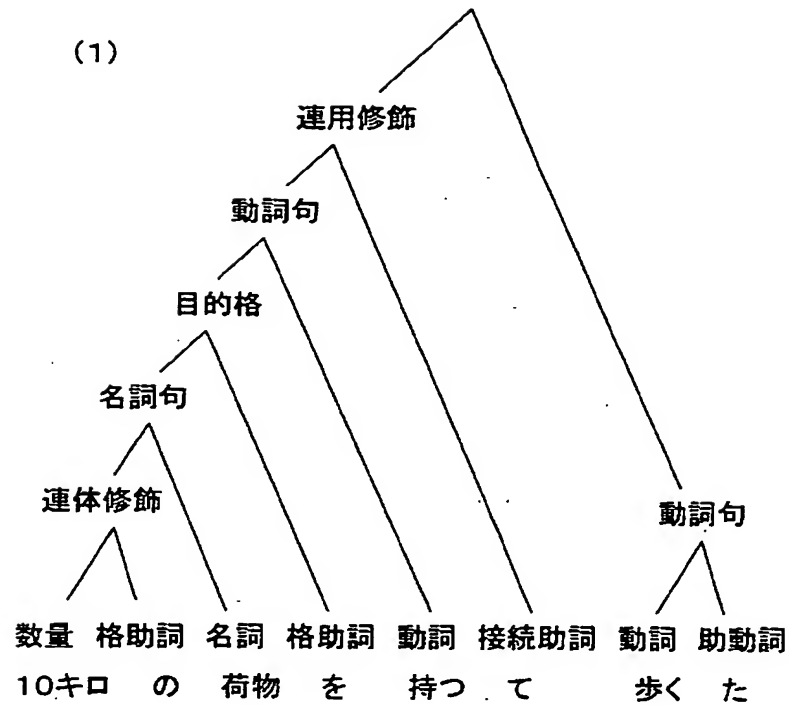
全体の流れ

【図 6】

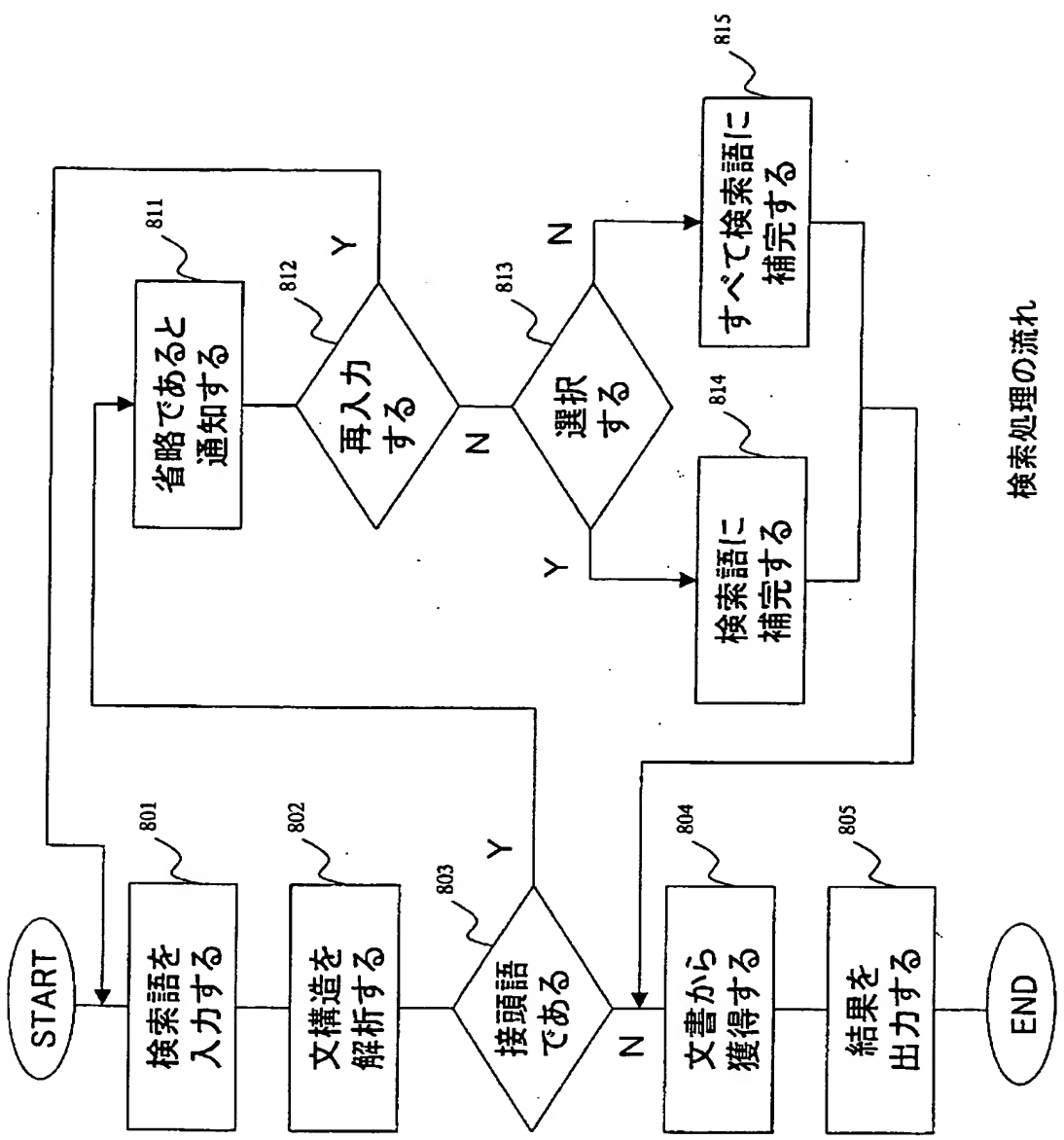


投入処理の流れ

【図 7】



【図 8】



検索処理の流れ

【書類名】 要約書

【要約】

【課題】 接頭語にみに省略されている場合を意識しなくても数量表現を検索することができる数量表現検索装置を提供する。

【解決手段】 検索対象の文書や検索したい数量表現を入力する入力手段 1 と、上記入力された文書や数量表現の文の構造を解析する文構造解析手段 2 と、属性を表す属性名と属性の意味を表す属性内容と補完するための基本単位とからなる属性情報、および省略されたと判断するための接頭語と接頭語の意味を表す倍数とからなる単位系情報が格納された属性辞書 4 と、属性を表す属性名と属性名を判断するための共起語とからなる情報が格納された共起語辞書 5 と、上記入力された文書や数量表現に対して、上記解析された文構造と属性辞書 4 を参照してまたはさらに共起語辞書 5 も参照して、上記文書や数量表現の接頭語に基本単位を補完することによって、省略された数量表現を補完する省略補完手段 3 とを備える。

【選択図】 図 1

特願 2 0 0 2 - 3 4 5 0 6 0

出 願 人 履 歷 情 報

識別番号

[0 0 0 0 0 0 2 9 5]

1 . 変更年月日

1 9 9 0 年 8 月 2 2 日

[変更理由]

新規登録

住 所

東京都港区虎ノ門1丁目7番12号

氏 名

沖電気工業株式会社